

NAMA : DITO DWI PRASETYO

KELAS : 3J TEKNIK INFORMATIKA

NPM : 2213020043

```
[1] import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import missingno as msno
```

```
[9] df = pd.read_csv('Titanic-Dataset.csv') #Ambil Data
df.shape #Menampilkan Ukuran Data

df.head() #Menampilkan 5 Baris Data
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

```
[10] print("\n--- Informasi Data ---")
df.info() #Informasi Dari Dataset
```

```
--- Informasi Data ---
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   PassengerId      891 non-null    int64
1   Survived         891 non-null    int64
2   Pclass           891 non-null    int64
3   Name             891 non-null    object
4   Sex              891 non-null    object
5   Age              714 non-null    float64
6   SibSp            891 non-null    int64
7   Parch            891 non-null    int64
8   Ticket           891 non-null    object
9   Fare             891 non-null    float64
10  Cabin            204 non-null    object
11  Embarked         889 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

```
[11] print("\n--- Missing Values ---")
      print(df.isnull().sum()) #Menampilkan Jumlah Missing Values
```

```

--- Missing Values ---
PassengerId    0
Survived        0
Pclass          0
Name            0
Sex             0
Age           177
SibSp           0
Parch           0
Ticket          0
Fare            0
Cabin          687
Embarked        2
dtype: int64

```

```
[12] #Visualisasi Missing Values
      plt.figure(figsize=(10,5))
      msno.matrix(df)
      plt.title("Visualisasi Missing Values")
      plt.show()
```



Berdasarkan dari dataset “Titanic-Dataset.csv” terdapat 3 atribut yang memiliki missing value (data yang tidak memiliki nilai / kosong), sebagai berikut :

1. Age (177 missing values)

- Tipe Data : Numerik (usia penumpang).
- Cara Menanganinya :
 - Mengisi dengan median usia agar tidak terlalu terpengaruh oleh outlier.
 - Mengelompokkan berdasarkan kelas tiket (Pclass) dan mengisi dengan median usia dari tiap kelas (karena usia dan kelas memiliki korelasi).
- Alasannya data tidak diisi oleh penumpang atau tidak dicatat dengan baik.

2. Cabin (687 missing values)

- Tipe Data : Kategorikal (nomor kabin).
- Cara Menanganinya :
 - Dikarenakan terlalu banyak data yang hilang, lebih baik menghapus kolom ini jika tidak terlalu penting.
 - Mengelompokkan berdasarkan kelas tiket (Pclass) dan memberikan label kategori seperti "Unknown" untuk penumpang tanpa kabin.
- Alasannya banyak penumpang tidak memiliki informasi kabin, terutama dari kelas ekonomi.

3. Embarked (2 missing values)

- Tipe Data : Kategorikal (pelabuhan tempat naik : C = Cherbourg, Q = Queenstown, S = Southampton).
- Cara Menangani :
 - Karena hanya 2 data yang hilang, bisa diisi dengan modus (nilai yang paling sering muncul), yaitu Southampton (S).
- Alasannya Data tidak tercatat atau hilang.